DOCUMENT RESUME

ED 068 517 TM 001 878

AUTHOR	O'Reilly, Robert P.; Gorth, William P.
TITLE	Alternatives to Accountability: Stool Pigeon Versus Servant and Soulmate.
PUB DATE	Oct 71
NOTE	22p.; Paper prepared for the Annual Conference of the Educational Records Bureau (New York, Oct. 28-29, 1971)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Academic Achievement; Behavioral Objectives; Computer Programs; Curriculum Development; Data Analysis; Data Bases; Decision Making; *Educational Accountability; *Evaluation Methods; *Instructional Systems; Management Information Systems; Scheduling; Speeches; *Teacher Responsibility; Test Construction; Test Results; Tests
IDENTIFIERS	Comprehensive Achievement Monitoring

ABSTRACT

The visible claims of the current accountability movement are examined critically, and an alternative philosophy and a developing system are offered. Areas examined include the psychological implications of accountability philosophies for teaching staff, certain educational measurement problems, and the availability or adequacy of operating systems required to support broad applications of accountability. The alternative philosophy offered focuses on operations for which teachers may legitimately be held responsible, as opposed to the current movement's focus on student outcomes. The comprehensive supporting information system is presented in broad outlines, highlighted by a presentation of an evaluative component called Comprehensive Achievement Monitoring. (Author/LH)

ED 068517

ALTERNATIVES TO ACCOUNTABILITY: STOOL PIGEON VERSUS SERVANT AND SOULMATE

Robert P. O'Reilly, Ph.D.
New York State Education Department

William P. Gorth, Ph.D.
University of Massachusetts

TM 001 878

October 1971

Prepared for the Annual Conference of the Educational Records Bureau, "Accountability: Its Impact on the Schools," held at the New York Hilton, New York City, October 28-29, 1971.

Schutz (1971) somewhat facetiously, labeled roles of measurement in education as including ". . . servant, soulmate, stool pigeon, statesman [and] scapegoat . . ." Each measurement era, it seems, emphasizes one or more of these roles to the relative exclusion of the others. Our observations of the current measurement era suggest that actors in the roles of stool pigeon and scapegoat are receiving the best reviews, while those in the more constructive roles of servant and soulmate are waiting in the wings. Sadly enough, this wait may be extended as the schools become the laboratory for a movement which again places the cart before the horse. We refer, of course, to the contemporary movement toward accountability whose unconstructive connotations may be partially communicated by its synonyms: reckoning, tidings, charge, score, inventory, report, and consequence. Judging from our ongoing efforts to establish performanced-based management information systems in the schools, it is the unconstructive connotations that teachers and others now associate with accountability in education.

Advocates of educational accountability, on the other hand, suggest that the movement will systematically aid educators to improve student achievement and effect needed changes in the schools (Lessinger, 1970). What can hardly yet be called a system appears to have as its prime elements: (a) evaluation procedures intended to measure growth in relation to desired outcomes of instruction, usually involving the use of standardized tests; (b) some form of public reporting of the extent to which desired outcomes were achieved; and (c) rewards to persons involved in the educational process, contingent upon the extent to which desired

outcomes are achieved. These elements may be augmented by additions to the school program to be subjected to test, and internal or external contractual agreements for program evaluation and/or the management of instruction.

Though recognizing the constructive spirit of accountability, we would nevertheless argue that current translations of the concept into operations (e.g., Dyer, 1970) and rational constructions (e.g., Barro, 1970) fail to take into account major contingent problems, which may prevent the movement from achieving any practical results in the contemporary scene. A first set of problems relates to the recognition of the psychological implications of the language and nature of the accountability movement for teachers and other professional staff. Evaluation procedures, for example, are explicitly external, (Martin, and Blaschke, 1971; Lessinger, 1970); the criteria for evaluation are further likely to be selected or designed by staff external to the school (Lessinger, 1970), and there is clear intent to report discrepancy data to such audiences as parents and boards of education. The most dense clinician would recognize in this set of events a clear cut basis for anxiety induction in the educational community, and would expect the usual chain of psychological reactions, ranging from expressed hostility to organized defensive responses designed to remove or afford escape from the offending stimulus. More extreme end results may include a temporary increase in the local base rate of paranoia, where accountability applications turn the educational and surrounding community into camps of accounters and accountants.

A second set of problems arises when the discrepant data are reported

to the public because the accountability philosophy may easily erode any remaining local confidence in the schools. This is particularly likely where an irresponsible press interprets test data only dimly understood by the public in the first place, and where there are organized community groups whose fixed orientation is to negatively reinforce the educational staff.

A third set of problems for the accountability philosophy relates to the professional implications resulting from externally imposed additions or changes to instructional programs. If the negative stimulus value of public disclosure of student performance data and "objective" staff evaluation are not sufficient to motivate instructional staff, then surely they will react defensively to outsiders assuming instructional and program responsibilities. On this matter, one may seriously question the generalizability of the notion that any external agencies can reliably produce meaningfully greater levels of achievement among children, given the results of large numbers of special efforts to date (Jenson, 1969; O'Reilly, 1970) and the quality of instructional programs and materials available to agents of accountability.

A fourth set of problems relates to the outcome measures proposed for use in accountability applications, typically standardized achievement tests. Timid criticisms suggest such tests are not entirely adequate as outcome measures, and must be supplemented by tests developed to measure other important outcomes (Lessinger, 1970). Other critics flatly conclude that standardized testing is irrelevant to the major questions asked in the accountability context, including those questions asked about

achievement of basic skills (Skager, 1971). On a yet more basic plane, one may convincingly argue that valid definitions of outcomes of school performance are not yet generally available and that, therefore, no set of adequate outcome measures exists. This last argument is convincingly developed by Rohwer (1971) in a discussion of contemporary procedures used to generate the objectives of formal schooling.

A fifth and final problem area relates to current capabilities to develop and operationalize relevant information systems and necessary data logistics and analytical systems required to answer primary questions of the accountability movement. Potentially adequate information systems are only now being tried on a small scale (Barro, 1970; Schutz, 1971). Experience further shows the operations of such systems are exceedingly complex, and greatly dependent on the active support of teachers and other staff. Clearly, considerable research, development, and installation activity is yet required before generally applicable systems are available in support of accountability. Such information system applications are also likely to be shortlived or superficial if the data generated do not relate functionally to the short- and long-term information needs of such audiences as teachers, supporting staff, and students.

The foregoing discussion presents a few supporting arguments for the simple proposition that engineering in education should be accomplished in full recognition of the nature and condition of the patient. The flavor of the accountability movement poses a new and seemingly powerful threat to the educational system. It is further a pretentious movement which assumes that precise causal effects can now be clearly demonstrated

in education, that these effects adequately define current values and needs, and that, therefore, related rewards and punishments can be established. Carried to extremes, the movement may be utterly demoralizing to those who assume close responsibility for the educational process by making questionable inroads into the functions that school personnel now regard as theirs (Cf. Mecklenburger and Wilson, 1971, 1971a). The ideals of responsiveness, relevance, and responsibility in education may, however, yet be advanced given appropriate recognition of the organismic character of the system. The remainder of the discussion will focus in part on the philosophical basis of a system which embodies the inherently constructive spirit of accountability, but avoids many of the problems apparent in publicized notions which now communicate hostile intent to the teacher in the field. A major part of the discussion is given over to a description of the nature of the system, its operation, and the assistance it affords the educational process.

Alternatives to Accountability

Language of the System

In the system to be described, the usual language of accountability is eschewed in favor of a more palatable notion that assumes responsibility of teachers and others for acting upon information made available on the performance of certain job-related operations. Working with teachers, curriculum coordinators, and school administrators, we first identify educational decision-making domains for which supporting information can

be made available through computer based systems. Sample decision-making domains and related responsible agents are summarized in Table 1.

Insert Table 1 about here

The first decision-making domain in Table 1 relates to the problems of defining and validating the behavioral outcomes of education, otherwise known as curricula. Currently, we show teachers how to proceed from a formal analysis of local needs through the establishment of the four levels of objectives defined in Table 2. All levels of objectives shown are behavioral, and constitute, respectively, the standards of operation

---

Insert Table 2 about here

---

for a program, a level within a program (both approximations and course objectives), and the series of behaviors which lead to the performance of complex operations, such as those defined by course objectives. For the past year, we have been engaged in the empirical derivation of a set of terminal objectives in reading, based on analyses of roles and reading performance requirements in actual life situations. Alternatively, for the other areas of the curriculum, teachers involved in our efforts attempt to establish their own terminal objectives on the basis of what they and others know of actual life requirements.

In the domain of curriculum development, the appropriate agents assume the responsibility of first defining tentative curriculum standards in the

form of the levels of objectives established for the sytem. From there, information systems based on criterion-referenced testing (CRT) (see Gorth, Schriber, and O'Reilly, 1971) deliver data relevant to the task of constructing adequate curricula. The responsible agents then examine data in relation to such curriculum issues as determining the relevance of objectives, establishing more appropriate sequences of objectives, and defining realistic levels of performance.

In the second domain in Table 1, the responsible agents initially assume the tasks involved in the design or selection of the materials and procedures which constitute a program of instruction. Appropriate CRT based systems then deliver information at the course objective level and above. Actions are then taken to continually adjust methods and materials to the point where the standards specified by course and terminal objectives are approximated by group performance data. Decisions in this area involve the replacement of materials and approaches, the appropriate placement of review, the addition of needed branches to the program, and so on.

In the third domain in Table 1, responsibility is assumed for the daily management of the instructional process, with the goal being optimization of learning environments for individual students. Teacher decisions relate to such functions as: (a) placing the student within the curriculum at a given level; (b) assigning appropriate instruction; (c) tracking performance signals as the student proceeds through a given package of instruction; and (d) assigning appropriate forms of remediation, selecting enrichment activities, or selecting the next set of objectives.

Moving finally to the domain of comparative product evaluation, we face a set of decision alternatives which ultimately must result in the selection of the relatively better program. Though we have not generally worked at this level, we would presume that program managers at different levels would essentially wish to select the best program on the basis of performance and costs. Included in this area is another set of decision alternatives resulting from an evaluative stance regarding schools or districts as objects to be subjected to comparative evaluation. Given adequate sets of terminal performance objectives and related measurement procedures, this may prove a useful stance upon which to base the rational allocation of resources, such as those available for special programs.

SPPED: A Computer Managed Information and Resource System*

The final phase of this discussion will consider the measurement, evaluation, and resource systems used to support the decision-making domains given in Table 1. The broad outlines of a project called SPPED (System for Pupil and Program Evaluation and Development) will first be presented, followed by a brief discussion of the characteristics of a major evaluative component of the sytem known as CAM (Comprehensive Achievement Monitoring).

*The SPPED system is under development by the New York State Education Department, in cooperation with staff from the University of Massachusetts (Dr. William Gorth), Stanford University (Dr. Paul Pinsky) and State University of New York at Stony Brook (Dr. Shelley Harrison).

The SPPED Project

The basic structure of the SPPED project is shown in Figure 1 as a set of five interrelated parts, each with a computerized support component.

Place Figure 1 about here

The first component of the SPPED is the central BOIR (Bank of Objectives, Test Items, and Instructional Resources), which contains extensively classified material used for: (a) curriculum development; (b) test item access; (c) derivation of local paper-based or computerized BOIR's; and (d) program design and installation. The complexity of the BOIR may be judged to some extent from the set of reading objectives now being stored. Approximately 2500 basic or generic objectives are being filed for computer access, along with 1000 course objectives and 800 terminal and approximate objectives. Each set of objectives is classified along as many as 50 different dimensions, each with one or several levels. In addition, the "O-bank" is to include several lists of elements which will enable the user to fill out each objective to the desired level of specificity (e.g., vocabulary lists, word elements, phonetic elements, etc.).

The primary intent of the "O-bank" is to automate the task of curriculum development through the process of computerized selection and modification. Entering the system, the user specifies a number of classifiers for content, level, and other identifiers. He then receives

a set of general or generic objectives to which he adds or specifies lists of elements (e.g., a word list), and indicates the number of levels into which objectives are to be grouped (e.g., grade levels, modules). The resultant output is a structured curriculum with objectives and associated elements grouped into the desired number of levels. Objectives are further organized into subgroups within levels; for each subgroup, output includes one or more criterion objectives (course or summary objectives). Once this selection process is complete, the local reading curriculum is deliverable on hard copy, IBM cards, or magnetic tape. A computer program is also available to facilitate the processes of adding to the local bank, and editing and deleting material.

The remaining components of the BOIR serve to index items and specific instructional resources in relation to each objective. The utility programs for each banking function allows continual up-date and refinement. Two secondary programs, Test Scheduling (TS) and Test Construction (TC), operate on the contents of the I-Bank to generate, respectively, random or specified test forms for CAM and mastery testing (MAST-T), and the printing of test schedules employing random, stratified sampling for CAM testing. The TS and TC programs are not restricted to use of the BOIR as a data base.

The components of the SPPED described to this point serve school personnel as basic resources for such functions as curriculum development, test development, instructional materials selection, and test

scheduling. The associated computer programs are designed to eliminate much of the massive paper handling and other clerical work that are necessarily part of any sizeable systems approach to instruction. The BOIR, TC, and TS programs were designed after four years of developmental work with CAM, and were meant to speed the process of CAM implementation. The reason for the additional SPPED resources becomes more apparent in the next section which briefly presents the specifications for CAM testing.

Comprehensive Achievement Monitoring (CAM)

A detailed technical description of CAM and related implementation procedures is presented in Gorth, Schriber and O'Reilly (1970). All of the decision-making is made on the basis of criterion referenced test results. The CAM design includes the following components:

1. The definition of a curriculum with behavioral objectives;

2. The writing of test items to measure student performance on each objectives which are criterion-referenced test items;

3. The organization of a set of randomly parallel tests, where each test is made up of all or a sample of items measuring all the objectives in the curriculum and therefore represents item sampling;

4. The design of a longitudinal, usually every three or four weeks, schedule of test occasions throughout the course;

5. The analysis of the test data and the reporting of results by computer, usually within a couple of days;

6. The interpretation of the results by evaluators, teachers and students as a means for making better decisions about their instruction and curriculum, and

7. The modification of curriculum, instructional activities and the CAM design based upon the results.

The CAM methodology has been designed to work well with any grade level or curricular area. In fact, it has already been used successfully in more than 20 schools, with more than 15,000 participating students, and at grade levels from 3rd to 12th and in every academic subject area (Allen and Gorth, 1971).

Particularly important to the success of this evaluation technique is the use of the computer. It alleviates the frequently encountered bottlenecks of most evaluations, the analysis of data and the reporting of results. The CAM computer program allows extensive freedom in the design of evaluations, incorporating both longitudinal testing with item sampling and mastery testing to correspond with traditional or unusual course designs.

The information which is typically provided in the CAM system includes:

1. For individual students

   a. the total score on the current test and all previous tests, and

   b. information on the correctness of their response to each item corresponding to course objectives on the current test; and

2. for any subgroup of students and any set of questions after each test administration

   a. the achievement level on each objective, and

   b. achievement profiles which display graphically the level of achievement on all objectives on the previous test occasions.

The latter information is identifiable as pretest, posttest, and long-term retention data for the duration of the course.

The computer program allows students' achievement to be plotted on any given objective (or group of objectives) for the entire course. This plot, called an achievement profile, gives a graphic presentation of the changes in group achievement throughout the course. Achievement profiles are a unique type of information available from the CAM model.

Figure 2 presents hypothetical achievement profiles for five objectives from a course. Brief comments below the graph give possible interpretations. It is obvious that achievement profiles provide a wealth of information, at whatever point in the course they are calculated. On

---

Insert Figure 2 about here

---

the pretest in the foregoing example, all objectives except number 2 show achievement at the chance level, or about 20 percent (five-option multiple-choice items). Several decisions could have been made after test administration one: (a) objective 1 was not learned, reteach it in some other

way; and (b) objective 2 has tested high on both the pretest and test administration 1, suggesting it would be safe to skip instruction in this objective. After test administration 5, two other decisions might have been made: (a) achievement on objective 3 seems to be slipping and review is needed; and (b) objective 8 seems closely related to objective 5 and perhaps should be taught now instead of later. CAM, therefore, represents an application of criterion-referenced testing to program evaluation performed with longitudinal evaluation using item sampling.

Conclusion

This has been a brief foray into the characteristics of a developing system and underlying philosophy designed to enable a systems approach to instruction in the local school context. In contrast with a great many articles on systems and models of accountability, the efforts described here have been designed and partially or wholly implemented in full realization of local needs and local potential for change. The simple wisdom in the philosophical basis of our work emanates from experiences with hundreds of teachers over several years. Some contrasts between this developing philosophy and the characteristics of accountability philosophies now prevalent in the literature are summarized as follows:

1. The language and focus of the system is on assuming responsibility for acting on information to effect changes in curricula,

instruction, and the learning activities of individuals -- as opposed to a direct focus on outcomes.

2. Teachers and related staff assist in formulating the outcomes of schooling in contrast with the external imposition of performance standards.

3. Teachers and related staff have a direct hand in specifying the measurements to be used in relation to determining performance, in selecting options available from different evaluation systems, and even in the formulation of evaluation systems as they are developed.

4. Program and instructional decisions are primarily the responsibility of regular instructional staff; external agents assist and advise in this process.

5. Public reporting of performance data, only now being considered, is to be initially based on simple (+, -) accounting of skills mastered, with the intent of involving parents continuously and functionally in the instructional process. Comparative performance data on program, schools, and individuals are not projected for use on the local level.

Administrators, board members, and others interested in accountability applications would do well to give very careful consideration to the various interrelated philosophical, psychological, technical, and logistical issues involved. Perhaps an appropriate starting point is to generate a local philosophy of accountability -- one which would be generally shared by all major participant groups. Once the sensitive issues are more or less settled and organized into a set of acceptable guidelines for development, the local leadership will more than likely find a more than adequate basis for proceeding with the business of accountability.

Allen, D. W. and Gorth, W. P. Fourth Annual Report to the Charles F. Kettering Foundation, Amherst, Mass.: School of Education, University of Massachusetts, 1971.

Barro, S. M. An approach to developing accountability measures for the public schools. Phi Delta Kappan, 1970, 52, 196-205.

Dyer, H. S. Toward objective criteria of professional accountability in the schools of New York City. Phi Delta Kappan, 1970, 52, 206-211.

Gorth, W. P., Schriber, P., and O'Reilly, R. P. Comprehensive Achievement Monitoring. Amherst, Mass.: School of Education, University of Massachusetts, 1971.

Jensen, A. R. How much can we boost IQ and scholastic achievement. Harvard Educational Review, 1969 (Winter), 39, 1-123.

Lessinger, L. Every kid a winner: accountability in education. New York: Simon and Schuster, 1970.

Martin, R. & Blaschke, C. Contracting for educational reform. Phi Delta Kappan, 1971, 52, 403-405.

Mecklenberger, J. A. & Wilson, J. A. The performance contract in Gary. Phi Delta Kappan, 1971, 52, 402-406.

_____. Performance contracting in Cherry Creek. Phi Delta Kappan, 1971a, 53, 51-54.

O'Reilly, R. Racial and social class isolation in the schools: implications for educational policy and practice. Praeger, 1970.

Rohwer, W. D. Prime time for education: early childhood or adolescence. Harvard Educational Review, 1971, 41, 316-341.

Schutz, R. E. The role of measurement in education: servant, soulmate, stool pigeon, statesman, scapegoat. Journal of Educational Measurement, 1971, 8, 141-146.

Skager, R. The system for objective-based evaluation - reading, its components, intended form, and how it is being evaluated. Unpublished report, Center for the Study of Evaluation, University of California, Los Angeles, California. June 30, 1971.

Table 1

Selected Decision Making Domains for Which Test Data Are Made Available in CMI Systems

Domain	Primary Responsible Agent
1. Curriculum Development	Curriculum teams, administrators, parents
2. Product Refinement	Teachers, teacher teams, administrators
3. Instructional Management	Teachers, students, parents
4. Comparative Product Evaluation	Administrators, program managers

Table 2

Levels of Objectives as Related to Derivation

Levels of Objectives	Derivation (from)
Terminal Objective	Actual or intended life performance situations
Approximations to Terminal Objectives	Four successively less difficult levels derived from terminal objectives
Course Objective	A temporary "stand-in" for approximations; a criterion or course or summative objective
Enabling Objective	One of a series of lower level objectives derived from course objectives or approximations

Figure 1

Components of the SPPED Project

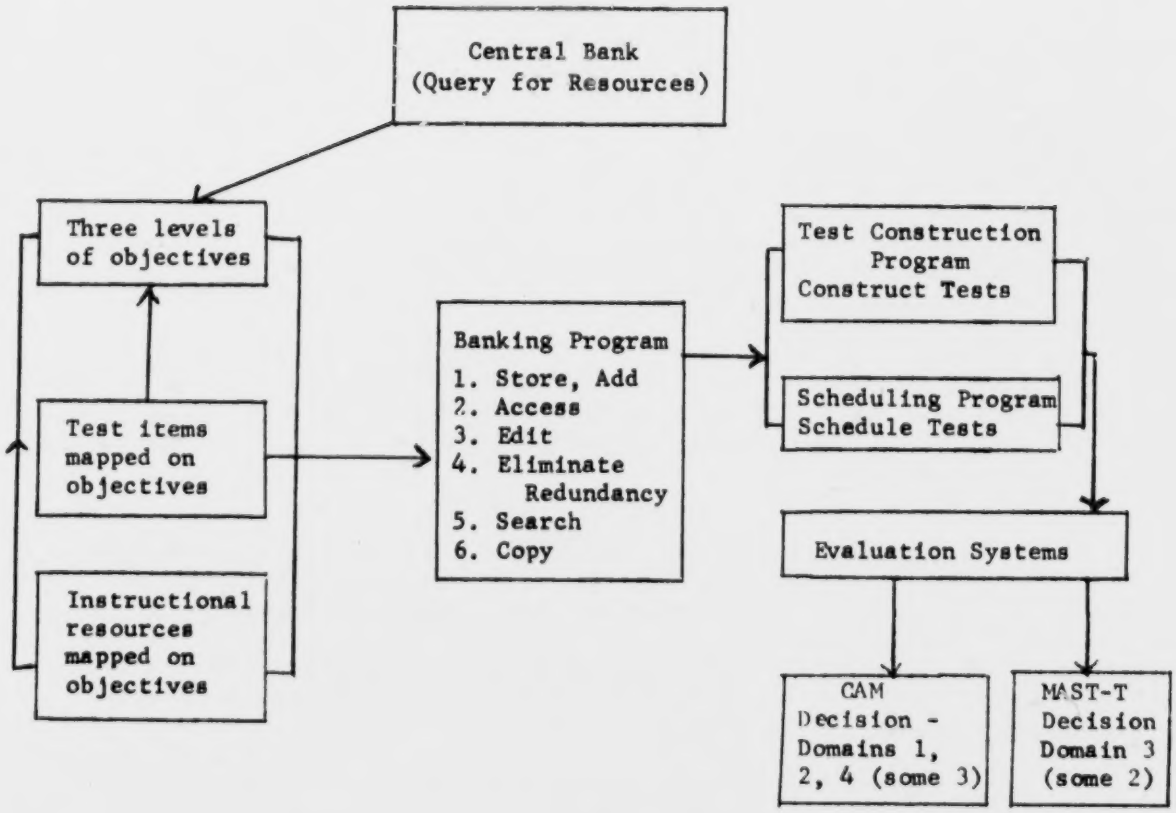
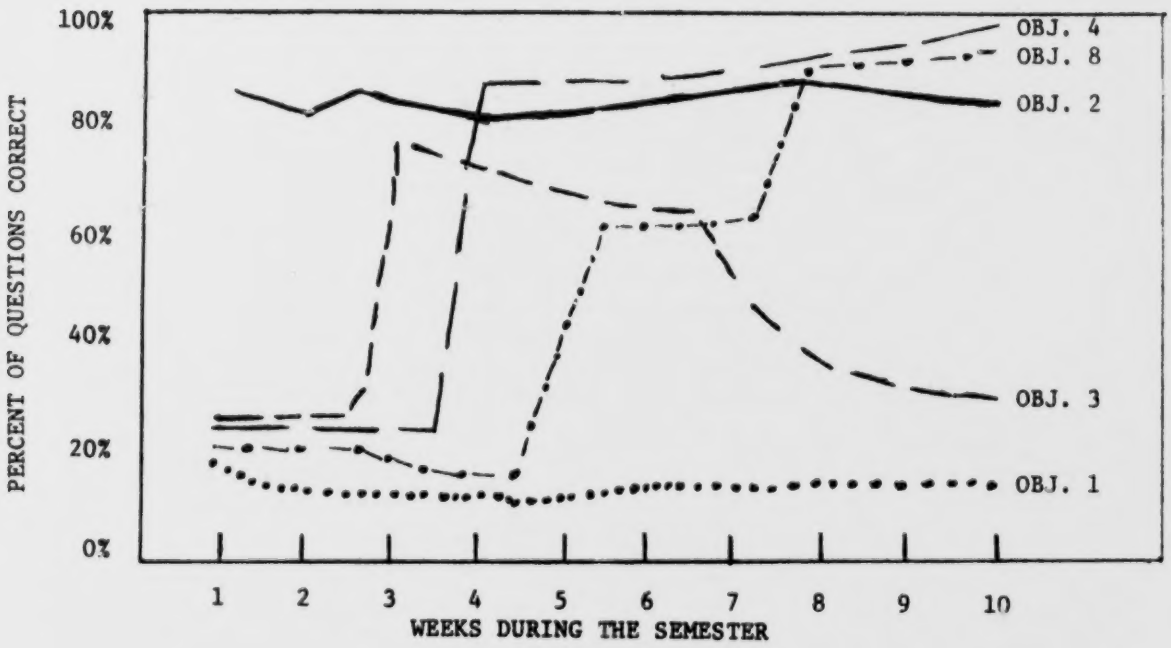


Figure 2

Achievement Profiles of a Group on Five Objectives



···· Objective 1: taught, but students did not learn; with rapid feedback, could be corrected with change in instruction (taught just before week 1)

—— Objective 2: previously known and not taught; without pre-test, this looks like student learning (taught just before week 2)

– – – Objective 3: taught and learned, but forgotten (taught just before week 3)

— — Objective 4: well taught (taught just before week 4)

—·— Objective 8: appears related to objective 5, because achievement increases when 5 is taught (taught just before week 8)

Abstract

This paper critically examines the visible claims of the current accountability movement and offers an alternative philosophy and a developing system. Areas examined include the psychological implications of accountability philosophies for teaching staff, certain educational measurement problems, and the availability or adequacy of operating systems required to support broad applications of accountability. The alternative philosophy offered focuses on operations for which teachers may legitimately be held responsible, as opposed to the current movement's overriding focus on student outcomes. The comprehensive supporting information system is presented in broad outlines, highlighted by a presentation of an evaluative component called Comprehensive Achievement Monitoring.

Authors

Robert P. O'Reilly, Ph.D., Cornell University. Now Chief, Bureau of School and Cultural Research, New York State Education Department. Direct development of the SPPED project, authored more than 40 papers and presentations on evaluation and measurement systems in education, desegregation research, and the measurement of anxiety in children.

William P. Gorth, Ph.D. Stanford University. Now Assistant Professor, School of Education, University of Massachusetts. Originator of CAM. Author of more than 70 papers, primarily on CAM, evaluation and measurement in education.